Statistical analysis of soil data from the McMurdo Dry Valleys, Antarctica

Linda Lilburne
GCAS 2004/5 Supervised Project

Abstract

Pedological data from over a decade of field trips to the McMurdo Dry valleys has been collated into a dataset. This data includes site observations (location, topographical position, estimated glacial history and soil age), morphological observations (full pit profile description), soil taxonomic classification, surface observations (weathering characteristics of boulders), chemical measurements of the major anions (Cl, SO_4, NO_4) and cations (Ma, Mg, Ca, K), electrical conductivity and pH.

An exploratory statistical analysis was performed on this dataset to determine which analyses were appropriate given the format of the dataset, and its quality and quantity. Box plots were used to study the variability of variables according to different groupings. Multivariate analyses including a factor analysis, discriminant analysis, cluster analysis and machine learning algorithms were all applied. Geostatistical analyses investigated the spatial dependence of some of the observations.

Most of the variability analyses indicated little differences in the ranges of soil properties between groups (weather stage, eco-climatic zone, taxonomic class, geological age). Where there were differences some trends were obvious and others were unexpected. The multivariate analyses did separate the pits and observations into groups that seem reasonably sensible. Little spatial dependence was found. It is concluded that the Bockheim dataset is sufficiently comprehensive for statistical analyses. The next stage in this work requires pedological input to refine those analyses that either have results of interest or have the potential to provide information of interest.

Introduction

The five factors that control soil development are recognised as being lithology, climate, topography, biological organisms and time. Due to extreme conditions such as lack of moisture and cold, there is very little biological activity in Antarctica, however, soils are still formed by the other four factors in the ice free regions (Tedrow and Ugolini, 1966). These ice free areas comprise just 0.3% of the Antarctic continent, $1/8^{th}$ of which is in the McMurdo Dry Valleys (New Zealand Antarctic Institute, 2001).

Soils have been systematically studied in Antarctica since the International Geophysics Year in 1957. While much of this research has focussed on contamination by hydrocarbons, some effort has also gone into pedological studies in particular by Campbell, Claridge and Bockheim. Antarctic soils are now referred to as cold desert soils, and are in US Soil taxonomic terms classified as Gelisols. They are characterised by very low temperatures and moisture levels, permafrost at varying depths, and a hard rocky or pebbly surface pavement (Campbell et al., 1998).

Pedological investigations over extensive areas usually involve digging pits at key sites to understand the soil development process. The number of pits is limited by the time it takes to access, dig and describe a soil, consequently there is usually little replication and random sampling techniques are rarely used. The season for digging in Antarctica is very short, and the logistical difficulties are considerable, so the number of sites is limited even further. Lack of independent data samples is a significant constraint to the application of statistical techniques.

Bockheim (2002) has put together a database of soil data collected over the years between 1975 to 1987. This project is an exploratory study to determine the suitability of the Bockheim dataset for statistical analysis. A preliminary analysis of the variability of key properties of interest, and any relationships between key morphological and chemical properties will be performed.

Background

A number of different soil-forming processes are evident in the McMurdo Dry Valley region. Older soils show more signs of weathering, both physical and chemical, whereas, more recent soils (< 50,000 years) show little soil development (Campbell et al., 1998). Soils may have been developing since the Early Miocene (Campbell and Claridge, 1987). Generally, soil cohesion, the depth and intensity of soil salinity and oxidation, and the fine earth fraction all increase with age. Younger soils are predominately sandy, whereas older soils have very small amounts of silt and clay (Bockheim, 1979). The oldest soils are found on surfaces furtherest from retreating glaciers, i.e., high upland areas.

A succession of glacial advances from both the east and west are a key influence on soil development. According to Hall and Denton (2005), drifts in the lower and central Wright Valley show that there were at least eight westward ice advances. Incursions (and the retreat) of ice and glaciers from the Ross Sea have a different effect on soil development to those from plateau or alpine glaciers. Also, because of different parent material composition, type of deposit, and weathering differences, soils formed on glacial deposits can vary significantly over short distances (Campbell and Claridge, 1987). Parent material in the Wright Valley is a mix of sandstone and dolerite. Glacier ablation and the expansion/contraction of patterned ground formations physically rework soil. There is only limited fluvial and aeolian transport of soil. In some places there is evidence of conveyer deposits (Hall and Denton, 2005).

Chemical weathering is controlled by the soil age, temperature and precipitation (Campbell et al., 1998). Older soils are more oxidised as indicated by deeper colouring and higher levels of free iron oxide. The chemical composition of salt accumulations in the soil, show their origin to be atmospheric deposition, rock weathering and marine. Often salts in a location may derive from all three sources (Campbell and Claridge, 1987). Near the coast salt has a marine origin and contains largely chloride and sulphate salts, whereas nitrate salts with little chloride are found further inland due to atmospheric deposition in snowfall (New Zealand Antarctic Institute, 2001).

Other important factors that result in variation of soil properties are variation in precipitation (wetter towards the coast), and differences in albedo (capacity to reflect solar radiation) as this impacts on soil temperatures and presence of summer meltwater (Campbell et al., 1998).

Methods

Data

A soil database produced by Bockheim and provided by Malcolm McLeod (2004) included eight ArcView shapefiles and three Excel spreadsheets (Table 1). Each site is numbered by year and order of visit. Data was collected from 1975 to 1987. The dataset is not complete in the sense that of the 479 pits recorded in *Bocheim desc with part-size.xls*, many do not have morphology, chemical or temperature data. All the shapefile attributes are described in Bockheim (2002). Table 1 indicates how many pits with data are in each file. Seven pits in *Bockheim desc with part-size.xls* have no horizon information at all. 187 sites in *S_SoilLocb.shp* are in the Wright valley; *V2 all bock WV data.xls* summarises most of the available attributes for Wright Valley pits.

Name	Description	Number with data	Attributes
S_soillocb.shp	Location of pits	392	
S_soilchemb.shp	Profile chemistry	113	Max electrical conductivity, salts to 70cm, major anions (Cl, SO_4, NO_3), % free iron, silt and clay
S_horbwchem.shp	Chemistry of the averaged Bw horizons	128	Top, Bottom, pH, electrical conductivity, major cations (Na, Ca, Mg, K), major anions (Cl, SO_4, NO_3), free iron, gravimetric moisture content.
S_horcnchem.shp	Chemistry of the averaged Cn horizons	96	Top, Bottom, pH, electrical conductivity, major cations (Na, Ca, Mg, K), major anions (Cl, SO_4, NO_3), free iron, gravimetric moisture content.
Bock □hem.. Data.dbf	Chemistry all horizon samples	394 samples, (less for NO_3)	Top, Bottom, pH, electrical conductivity, major cations (Na, Ca, Mg, K), major anions (Cl, SO_4, NO_3), free iron.
S_soilairtb.shp	Air and soil temperature	311 (air) 163 (soil)	Air temperature, Soil temperature in up to 5 horizons
S_grdiceb.shp	Ground ice features	111	Ground ice feature category
S_bldfeatb.shp	Boulder features	232	Boulder frequency, ratio of sandstone to dolerite, % with glacial striations, desert varnish, spalling, pitting and ventification
S_soilmorphb.shp	Soil morphology	381	Depth of staining, coherence, visible salt, salt pan, weathered clasts (ghosts), max colour development, and ice cemented layer, weathering stage, morphogenetic salt stage, permafrost type

V2 all bock WV data.xls	Wright Valley data	178	Relief, eco-climate, patterned, age, location, soil classification, texture, landform, parent material, ground ice, depth of staining, coherence, visible salt, salt pan, weathered clasts (ghosts), max colour development, and ice cemented layer, weathering stage, morphogenetic salt stage, boulder frequency, ratio of sandstone to dolerite, % with glacial striations, desert varnish, spalling, pitting and ventification, max electrical conductivity, salts to 70cm, major anions (Cl, SO_4, NO_3), % free iron, silt and clay
Bockheim desc with part-size.xls	Full soil description for all pits	480	Location, landform, parent material type and age, relief, ecoclimate, patterned, soil classification, and horizon data: notation, depth, boundary, Munsell colour, texture, structure, consistency, salt morphology stage, cementation, reaction to HCI, % stone, cobbles, gravel and coarse fragments

These files were reformatted as required into a form suitable for statistical analysis. This involved manual and automated manipulations of the *Bockheim desc with part-size.xls* spreadsheet. A visual basic program was written to facilitate the extraction of data from this spreadsheet.

Other layers include 50 m contours from Bockheim, surficial geology (based on Prentice polygons) and a satellite image of the Wright Valley. An overlay analysis with the surficial geology polygons provided an estimate of the glacial drift and age of each pit.

Statistical analyses

A wide variety of statistical techniques have been applied to pedological data, ranging from simple statistical descriptors of variability, correlation, analysis of variance (ANOVA), clustering techniques, factor analysis, regression, and spatial statistics. Each of these techniques will be applied or discussed in this report. Further details of these techniques can be found in Hair et al. (1995)) and Systat (1997).

The pit profiles have been classified according to the US Soil Taxonomy (Soil Survey Staff, 1999) as either anhyorthel gelisols or anhyturbel gelisols. Pits were further classified as one of typic, glacic, salic, nitric or petrosalic Weathering stage is an indicator of the age of the soil as described in Bockheim (1979). The variability of depth of ghosts, oxidation, cemented ice, visible salt and salt pan within the soil classes and ages were studied graphically with box-plots. Correlation was inspected by viewing a pairwise scatter plot of the relevant variables.

Factor analysis summarises the data covariance structure in a smaller number of dimensions. The emphasis in factor analysis is the identification of underlying "factors" that might explain the dimensions associated with large data variability. There are no particular statistical assumptions underlying factor analysis, indeed some degree of multicollinearity should exist so that factors can be identified. The principal components method with a varimax rotation was applied.

Discriminant analysis is useful in estimating relationships between independent numeric variables and a categorical variable. It is a way of understanding group differences.

Discriminant analysis is based on linear relationships, and assumes multivariate normality and equal covariances among the groups. A discriminant analysis was applied to see if the numeric soil properties were good descriptors of the US taxonomic class.

A cluster analysis is similar to discriminant analysis in that a set of objects is classified into groups except that the number and membership of the groups is unknown. There are different techniques and different metrics of similarity. A hierarchical agglomerative analysis is applied here using squared Euclidean distance and average linking. Cluster analysis does not rely on any distribution assumptions.

A machine learning tool called WEKA (Witten and Frank, 2000) includes a number of data mining algorithms, that were applied to the Wright Valley dataset by Brent Martin of Canterbury University. These include Ripple down rule learner (Ridor), Non-nested generalised exemplars (NNGE), J4.8, Naïve Bayes, K nearest neighbour (KNN).

Geostatistical analyses investigate spatial dependence or the auto correlation of sample data over space. One method is to inspect a semi-variogram. If there is spatial dependence the points will tend upwards to start with, then flatten out as distance increases. If there is no initial upward trend, then close-by pits are not similar.

Regression type techniques were not applied as the current focus on soil research in the Wright Valley is on understanding soil development processes and classification rather than on predicting unmeasured soil properties.

Results

A map taken from the McMurdo Dry Valleys ASMA No. 2 Management Plan (Figure 1) depicts the Taylor and Wright Valleys. The Wright Valley pits are shown on Figure 2 (the length of the valley is approximately 60 km). There are considerably more pits in the eastern or lower Wright Valley. Weathering stage varies throughout the valley with unweathered soils adjacent to pits showing maximum weathering. Other variables show similar spatial variability.

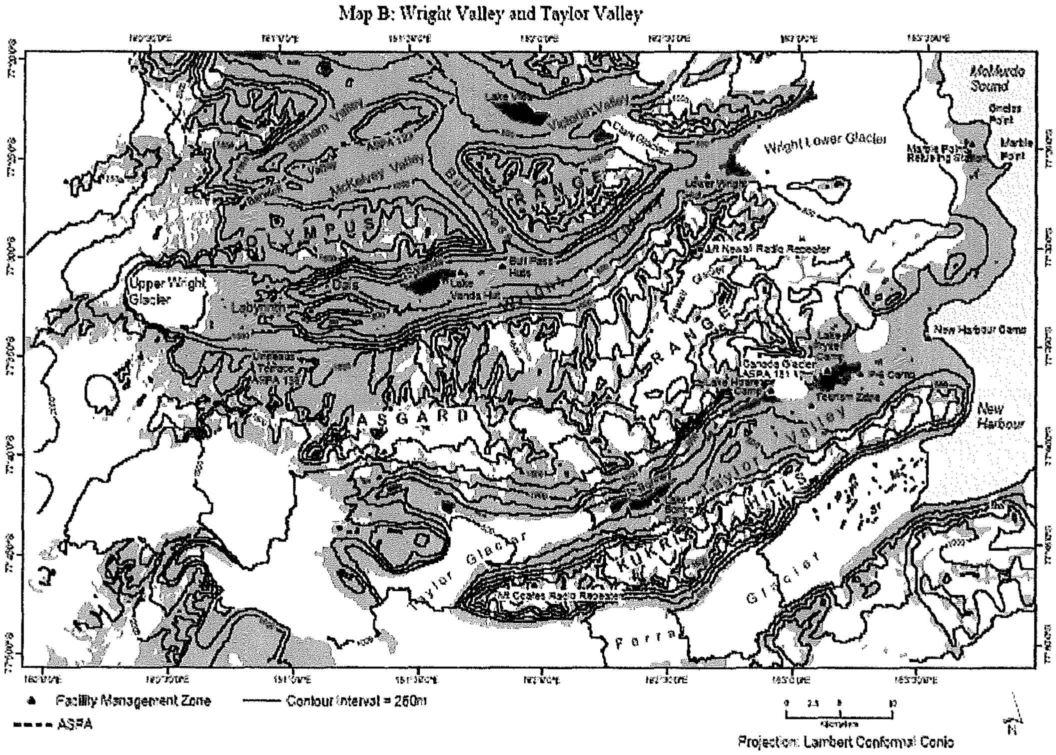

Figure 1 Map of the Wright and Taylor Valleys

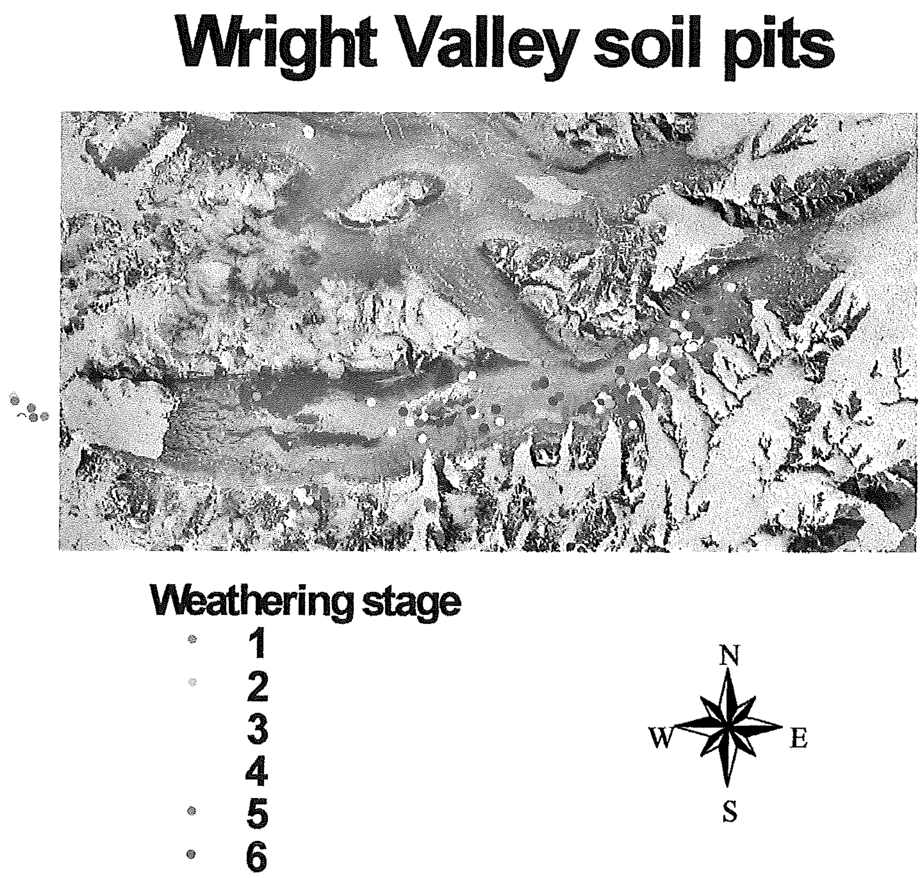

Figure 2 Soil pits in the Wright Valley. Coded by weathering stage (1 = minimal weathering, 6 = maximum weathering)

Table 1 shows the frequency of soil property observations in Wright Valley pits. Physical characteristics are almost complete, chemical measurements are available for only one third of the pits, and particle size and moisture measurements are only available for a few pits. Most of the pits are Typic Anhyorthels, with some Salic and Petrosalic Anhyorthels, and Typic Anhyturbels. There are a handful pits that are classified as other suborders (Nitric, Lithic, Glacic, Gypsic and PetroGypsic). Taxonomic codes are in Table 2. The pits are almost all sandy or sandy-skeletal.

Table 1. Frequency of observations for soil properties, of each taxonomic code, and of texture class for Wright Valley pits.

Soil property	**n**	**Tax Code**	**n**	**Texture**	**n**
Depth of staining	177	ABCA	1	Coarse silt	1
Depth of ghosts (boulders completely disintegrated to sand)	176	ABCB	7	Coarse silt/sand	1
Depth to Ice cemented layer	175	ABCC	1	Sand	28
Depth of coherence	177	ABCF	8	Sand-skeletal	144
Max CDE	175	ABCH	26	Sand-sketetal/sand	2
Salt stage	177	ACCA	2	Sand/sand-skeletal	1
Weathering stage	175	ACCD	2		
Max EC	77	ACCE	2		
Depth of visible salts	177	ACCF	21		
Salts to 70cm	77	ACCH	86		
Depth to salt pan		ACCI	21		
Chloride	70				
Sulphate	70				
Nitrate	70				
Free iron oxide	21				
Silt & Clay	29				
Boulder striations	54				
Boulder pitting, spalling, varnish and ventification	64				
Number of boulders	99				
Dolerite sandstone ratio	13				

Table 2. Codes of US taxonomic soil classes found in the Wright Valley by Bockheim.

US Taxonomic classification for Gelisols			
ABCA	Lithic Anhyturbel	ACCA	Lithic Anhyorthel
ABCB	Glacic Anhyturbel	ACCD	GypsicAnhyorthel
ABCC	Petrogypsic Anhyturbel	ACCE	Nitric Anhyorthel
ABCF	Salic Anhyturbel	ACCF	Salic Anhyorthel
ABCH	Typic Anhyturbel	ACCH	Typic Anhyorthel
		ACCI	Petrosalic Anhyorthel

As can be seen in Figure 3, soils representing the full range of weathering stage and salt morphology stage were found by Bockheim. As there were more sites closer to the coast, chloride levels were skewed to the higher end, and sulphate and nitrate skewed to the low end.

Figure 4 uses boxplots to show how soil properties vary by taxonomic class. There is little difference between either the turbels and othels or between subgroup level e.g., typics vs salics.

The ice cemented layer is quite shallow in lithic soils, as expected by definition. Some salt related properties do however show a difference between typic and salic soils (as expected). Salic and petrosalic soils have higher salt morphogenetic stage, max electrical conductivity, salts to 70 cm, depth of visible salts than do the typic soils. Outlier pit 84-96 may be misclassified in that it is a petrosalic with a salt stage of just 1.

A similar examination of boxplots (Figure 5) for the whole McMurdo Dry Valley dataset show a similar pattern, i.e.,

- glacics and lithics are shallow to ice cemented layer (<50cm)
- salics and petrosalics have higher levels of salts to 70 cm (4000+) and high electrical conductivity (>30dS/m) and a salt stage > 4. Depth of visible salts also tend to be higher.
- salics tend to be more weathered (≥4) whereas the other subgroups are mostly ≤4
- depth of ghosts, maxCDE varies with subgroup
- pits that are not typic, glacic or lithic tend to have higher depths of coherence (>50cm)

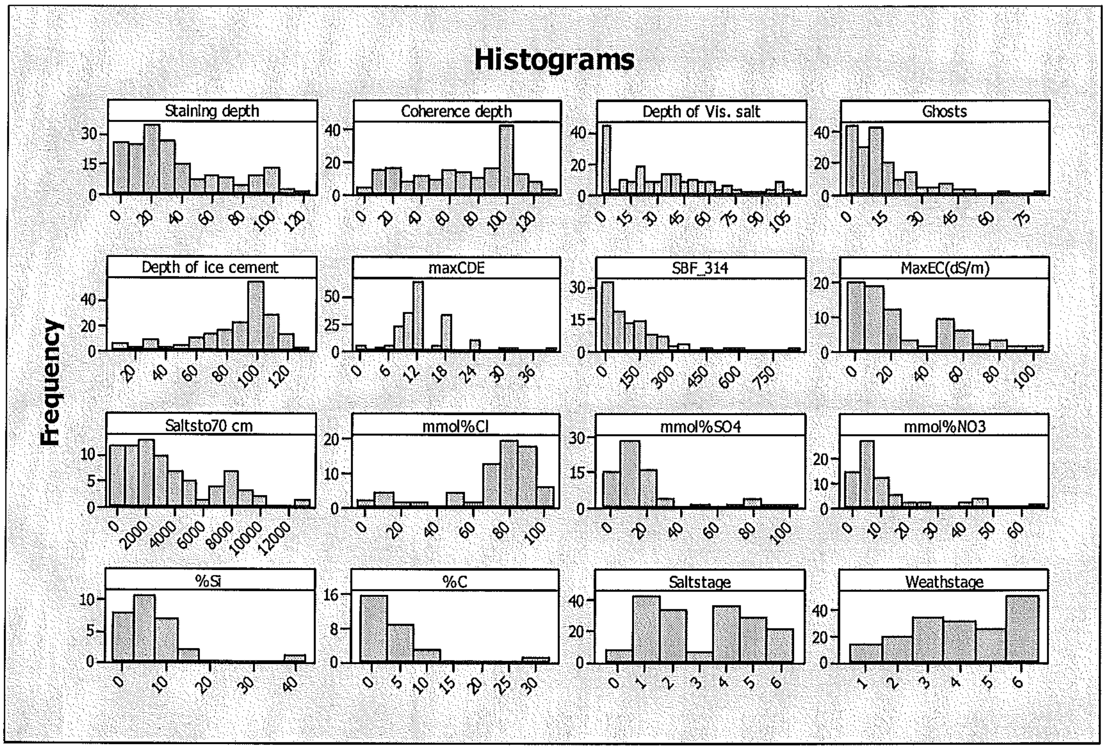

Figure 3 Histograms of key soil properties from Wright Valley pits

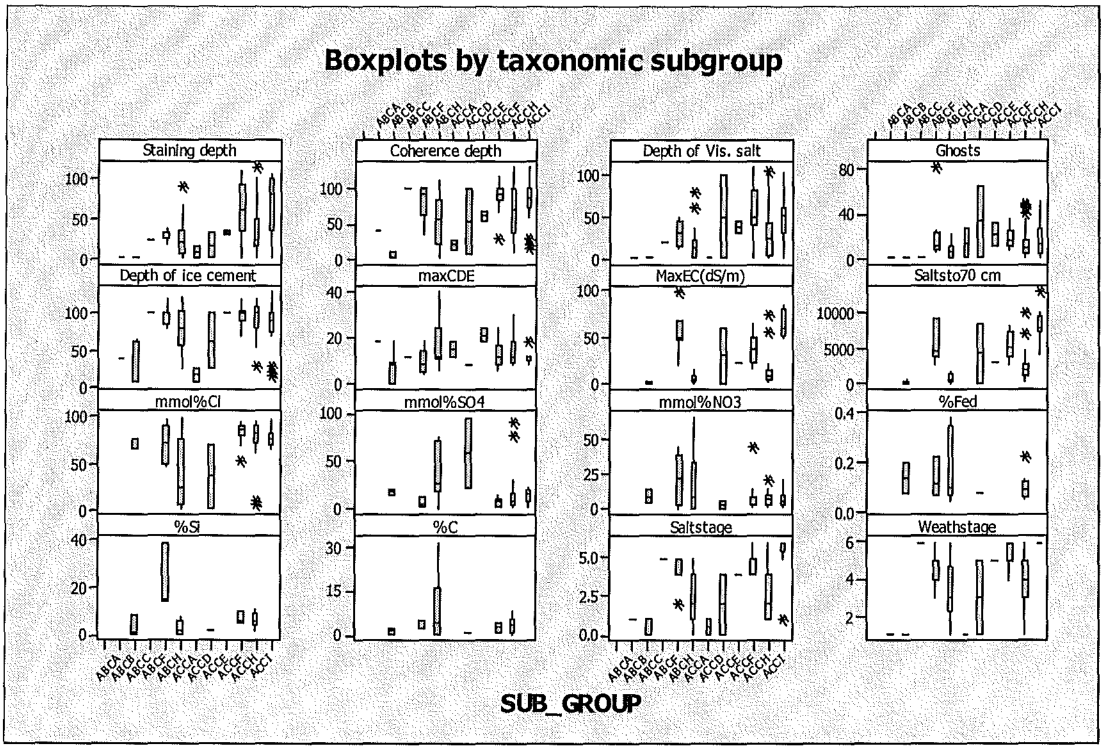

Figure 4 Boxplots of Wright Valley soil properties by taxonomic class

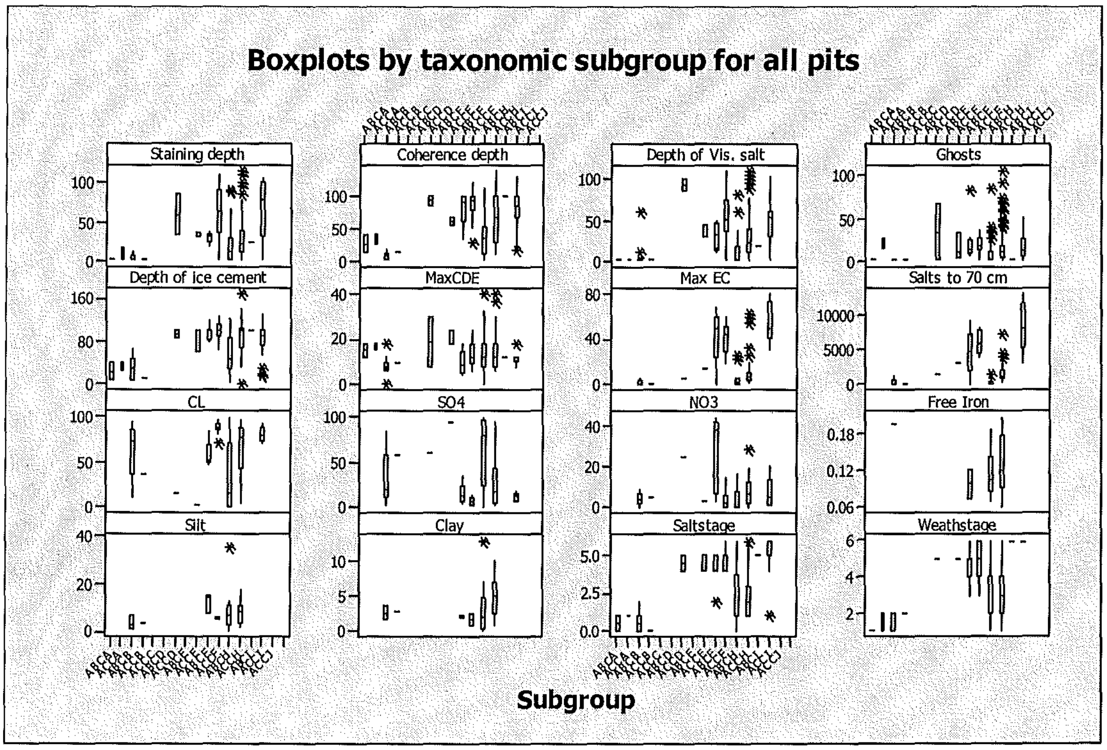

Figure 5 Boxplots of soil properties for all pits in the Dry Valleys

Boxplots (not shown) by eco-climatic zone show practically no differences in soil variability (except for more silt in the Coastal zone). Boxplots by weathering stage (Figure 6) show increasing ranges for depth of coherence and staining, EC and salts to 70 cm. Other soil properties did not vary (boulders, max CDE, ghosts, depth to ice cemented layer).

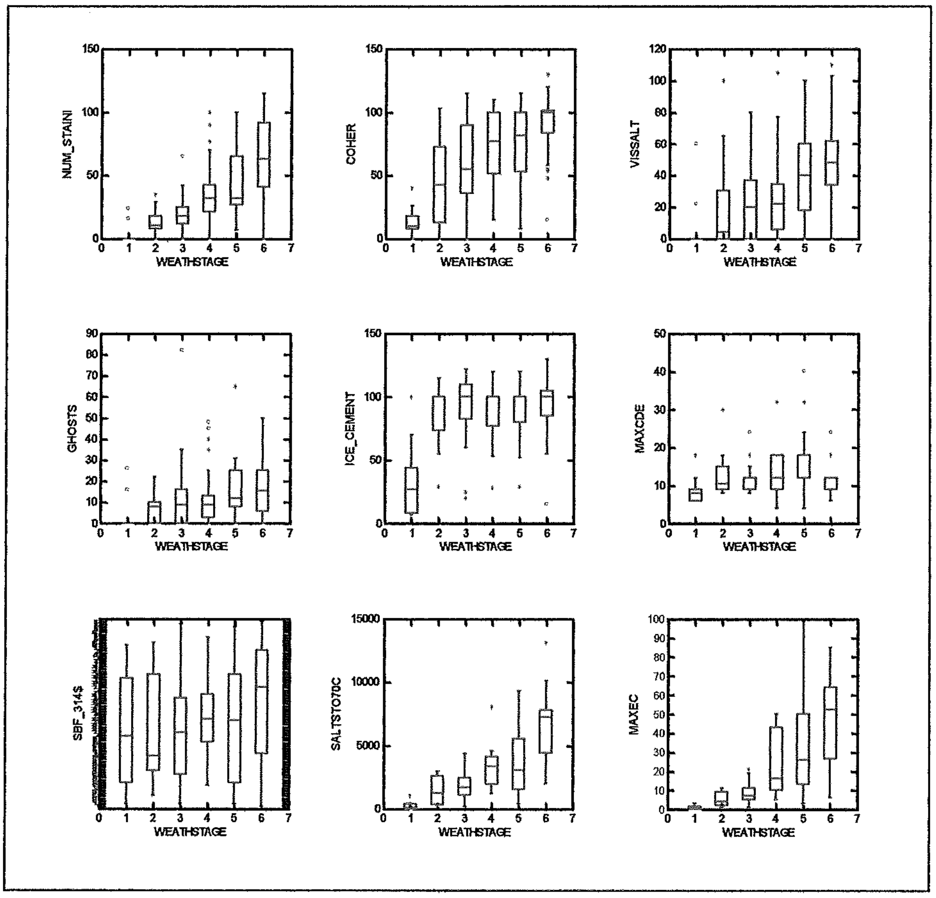

Figure 6 Box plots of Wright Valley soils by weathering stage

Pits of all weathering stages can be found in each of the eco-climatic zones. The following output shows that lithics are only found in the upland valleys, salics and petrosalics are mostly found in the inland valley (side or floor), and typics are found everywhere.

	Coastal	Inland valley floor	Inland valley side	Plateau fringe	Upland valley	All
ABCA	0	0	0	0	7	7
ACCA	0	0	0	0	3	3
ABCB	0	4	8	2	1	15
ACCB	0	0	1	0	0	1
ACCC	0	0	0	0	2	2
ABCD	0	0	0	0	2	2
ACCD	0	1	0	0	2	3
ABCE	0	0	0	0	2	2
ACCE	0	0	1	1	2	4
ABCF	0	6	2	0	0	8
ACCF	0	9	20	0	1	30
ABCH	18	31	20	2	36	107
ACCH	3	45	133	6	63	250
ABCI	0	1	0	0	1	2
ACCI	0	11	11	0	3	25
ACCJ	0	0	0	0	4	4
Missing	0	3	4	0	0	*
All	21	108	196	11	129	465

Figure 7 is a boxplot of all sites by geological age. Curiously, it shows depth of staining, visible salts and ice cemented layer to be least in the youngest and oldest periods and greatest in the Pliocene.

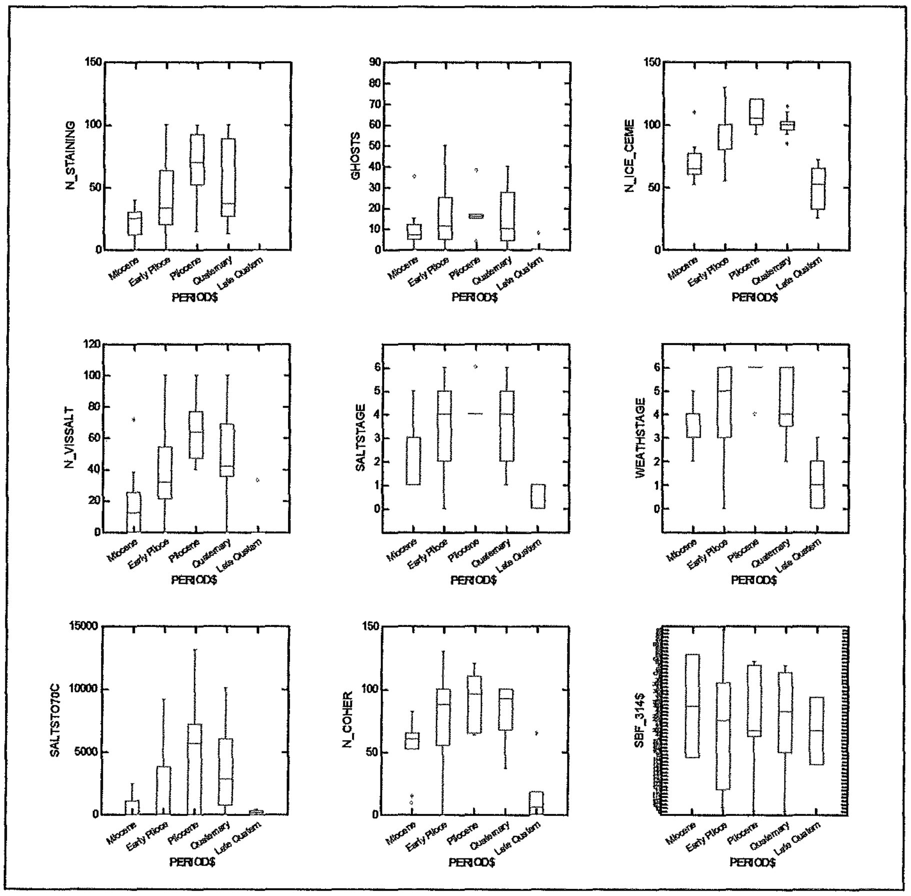

Figure 7 Boxplots of some soil properties of Wright Valley pits by geological age (as estimated by an overlay with the surficial polygons)

A pairwise scatterplot of all the numeric variables can quickly identify any correlated variables (Figure 8). The only pair that seem to be correlated are max EC and salts to 70cm which is not surprising given that one is calculated from the other (Bockheim, 2002).

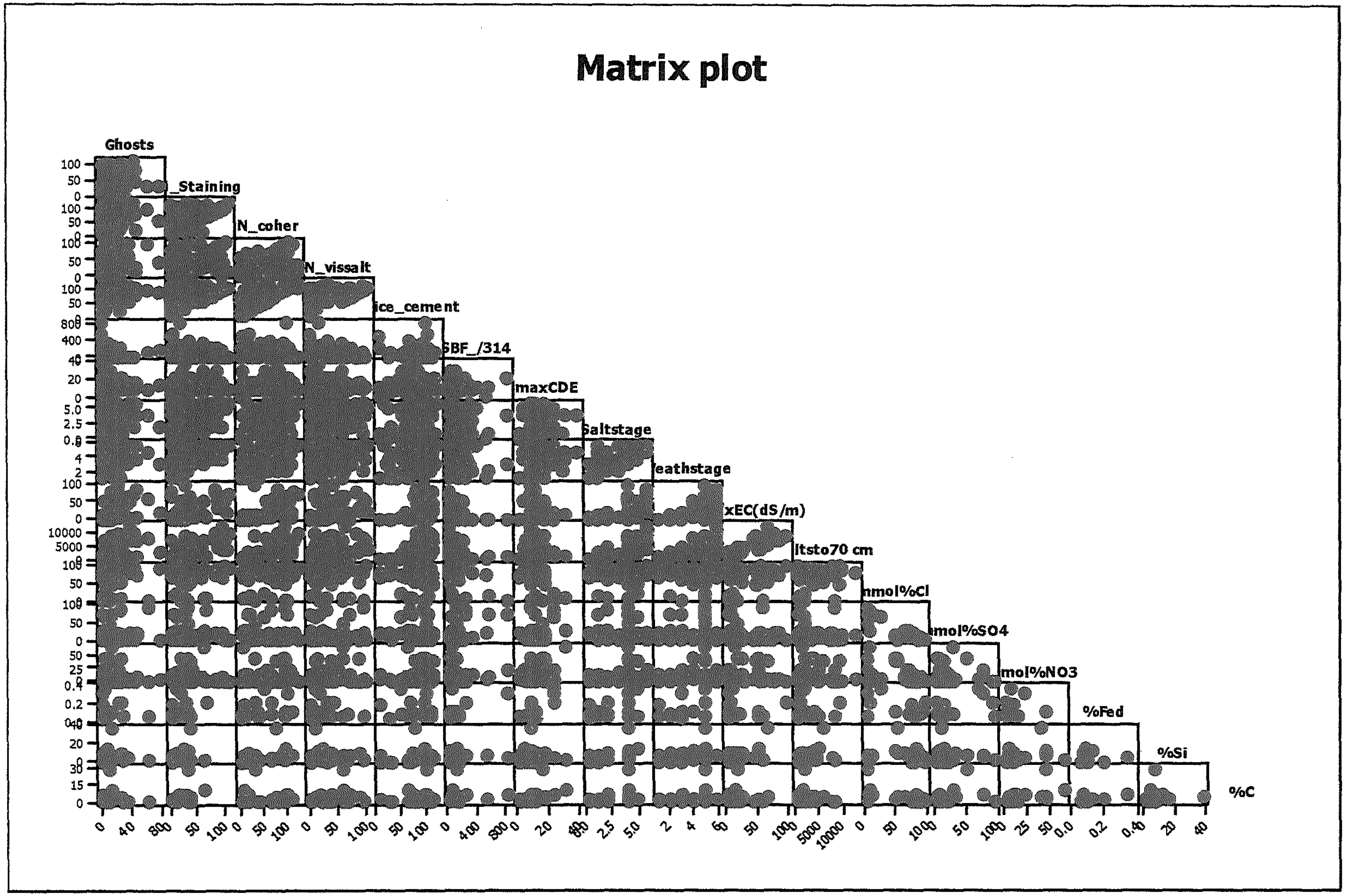

Figure 8 Pairwise scatter plots of the Wright Valley data

Factor analysis

A number of analyses were run with different sets of variables. Pits with any missing data are excluded so some variables are best dropped from the analysis e.g., silt, clay. The following output shows the loadings on four factors (which explain 83% of the variance). Factor 1 is a combination of weather stage, salt stage, depth of staining, depth of visible salt, salts to 70 cm and max EC. Factor 2 is related to max CDE, and to a lesser degree salts to 70 cm and max EC. Factor 3 relates to depth to ice cemented layer, and factor 4 relates to depth to ghosts.

```
Rotated Loading Matrix ( VARIMAX, Gamma =       1.0000)
```

	1	2	3	4
WEATHSTAGE	0.869	0.150	0.318	0.121
SALTSTAGE	0.853	0.103	0.358	0.119
NUM_STAINI	0.794	0.065	0.047	0.272
SALTSTO70C	0.753	-0.335	0.373	0.281
MAXEC	0.737	-0.389	0.338	0.249
VISSALT	0.704	0.252	0.075	0.088
MAXCDE	0.139	0.913	0.130	0.090
ICE_CEMENT	0.119	0.130	0.900	0.096
COHER	0.461	-0.003	0.725	0.108
GHOSTS	0.290	0.087	0.145	0.934

In another analysis where the salts were included, NO_3, SO_4 and Cl also loaded onto factor 2 (Figure 9).

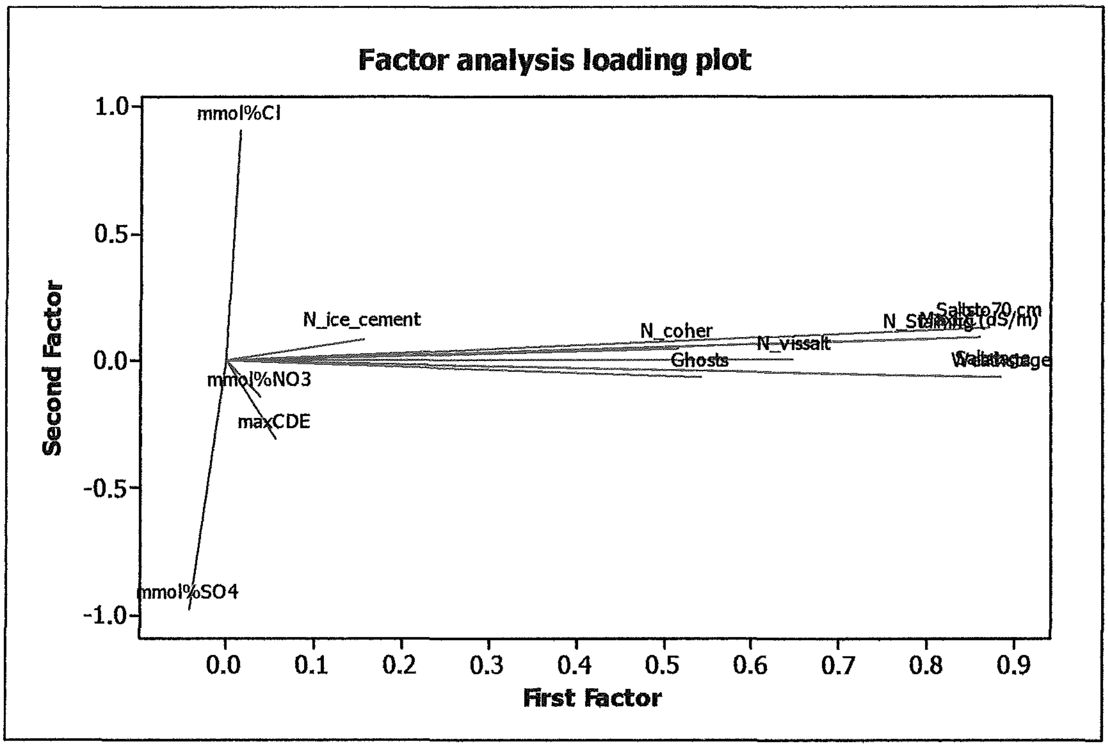

Figure 9 Factor analysis loading plot of the first two factors (Wright Valley dataset)

Given a focus on identifying theoretically meaningful factors (as opposed to reducing the number of variables) an oblique rotation may be more appropriate (Hair et al., 1995).

Discriminant analysis

The following variates were used to predict taxonomic class: depth to ice cemented layer, depth of staining and ghosts, max CDE, weather stage, salt stage, salts to 70cm and max EC.

Nitrate, sulphate and chloride were highly correlated with other predictors so were dropped. Discriminant analysis does not handle missing data, reducing the number of cases to 70. Taxonomic groups with hardly any cases were removed leaving ABCB, ABCF, ABCH, ACCF, ACCH, and ACCI.

Summary of classification

Put into Group	ABCB	ABCF	ABCH	ACCF	ACCH	ACCI
			True Group			
ABCB	4	0	3	0	0	0
ABCF	0	6	0	2	0	2
ABCH	0	0	4	0	5	0
ACCF	0	1	0	7	2	0
ACCH	0	0	3	1	23	0
ACCI	0	0	0	1	1	5
Total N	4	7	10	11	31	7
N correct	4	6	4	7	23	5
Proportion	1.000	0.857	0.400	0.636	0.742	0.714

N = 70 N Correct = 49 Proportion Correct = 0.700

70% were accurately classified, and if the turbel/orthel division (i.e., crytoturbated or not) is ignored 86% were correctly classified.

Some further work is needed to check the impact of violating the assumption of multivariate normality. As can be seen from Figure 3 some of the variates are not normally distributed. A square root transformation could be tried. A quadratic discriminant analysis would be more appropriate if covariances are inequal (Minitab, 2003).

Cluster analysis

Five clusters seemed appropriate given the small differences in similarity for six or more clusters. The resulting dendrogram (split over three panels) is shown in Figure 10. The third cluster has only one pit (78-06), which has very high depth to ghosts. The next cluster (4) has eight pits and is characterised by high max CDE. Cluster 1 is a group of 15 poorly developed soils with low depth of ghosts, coherence, and staining, low max CDE, shallow permafrost, low salt morphogenetic stage. Cluster 5 contains 12 well-developed pits with higher depths of staining, coherence and ghosts (excepting 78-06) and salt stage development. The final cluster (2) has 128 pits of average soil development and greater depth to the ice-cemented layer. Pedological input is needed to further understand the relevance of these clusters.

Cluster Centroids

Variable	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5
Num_Staining	-1.07428	-0.008848	-0.10939	0.28374	1.25718
Num_coher	-1.58219	0.169127	-0.35656	-0.24394	0.36606
Ghosts	-0.72518	-0.153035	5.37468	-0.06933	2.13718
Num_ice_cement	-2.33020	0.325116	0.08423	-0.83602	-0.00483
maxCDE	-0.72396	-0.107119	0.41260	2.82926	0.12700
Saltstage	-0.94213	0.047277	-0.58489	-0.11600	0.79945

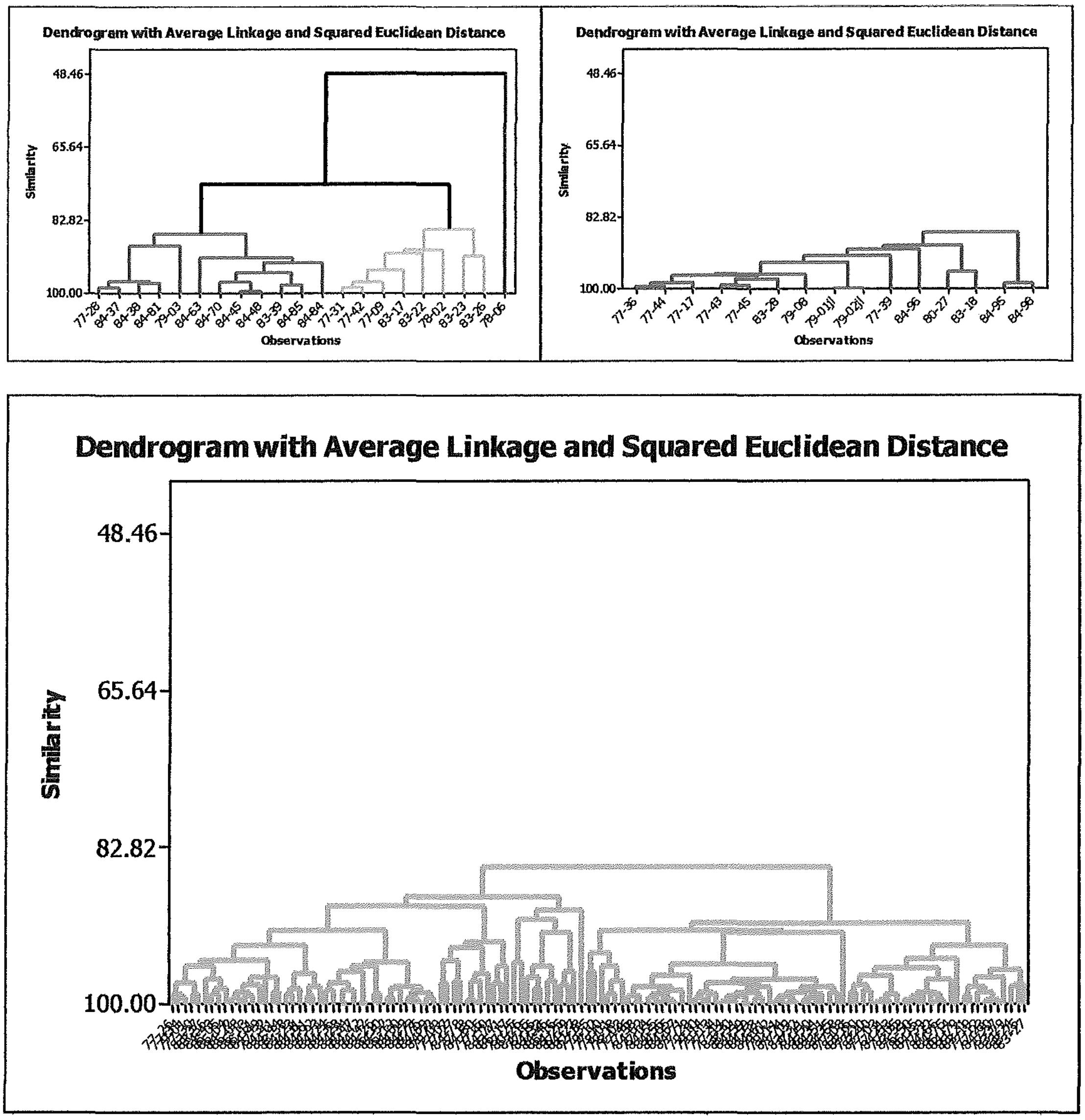

Figure 10 Dendrogram of a cluster analysis of Wright Valley pits

Machine learning analyses

Brent Martin's report is attached as Appendix Antarctica Accuracy of prediction of the taxonomic class was not as good as the discriminant analysis results. The Ridor algorithm was the best at 65% accuracy, however the classes with only a few cases were not removed making this a more challenging dataset than that used in the discriminant analysis. The petrosalic pits were identified by having a salt pan, as were the salics but these had a lower salt morphogenetic stage. The glacic pits were identified by a shallow depth to permafrost. The typic turbels were separated from the typic orthels as being shallower to permafrost. These derived rules really only verify the obvious rather than provide any interesting insight.

The J4.8 algorithm (a later version of C4.5 decision tree algorithm) generated more complex rules than the Ridor algorithm. It would be preferable for a pedologist to examine these rules for relevance. The salt pan switch and salt pan depth, which were not used in the other analyses, proved to be important nodes in the decision tree. The other algorithms did not appear to hold much promise for useful analysis of the Wright Valley dataset.

Geostatistics

ArcGIS was used to generate some semi-variograms of key soil properties (Figure 11) over a variety of lags. There was no indication of spatial dependence for weather class, salt stage class and depth to permafrost (despite the fitted yellow line). Chloride showed limited spatial dependence.

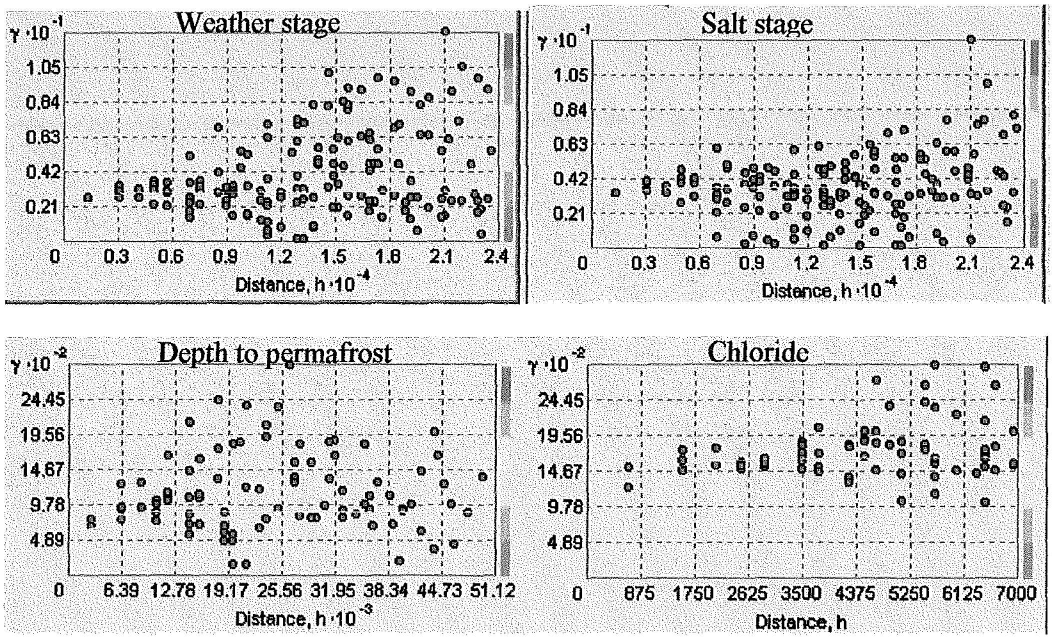

Figure 11 Semi-variograms of four soil properties from Wright Valley pits

Discussion

Data

Most of the depth variables had cases where a '>' or '<' was used, presumably to indicate approximate depth, or possibly unknown depth. In all cases the symbol was removed and the variable converted to a numeric form. This may result in misleading data. Salt pans were also difficult to handle as empty cells may indicate no salt pan or missing data. For the machine learning study, a switch was added to indicate whether a salt pan was present or not. Empty salt pan cells were assumed to mean no salt pan was present.

The complex multi-layered nature of glacial advances means that some pits have buried layers making it difficult to assign profile properties. Pit 84-96 has already been mentioned – it has a buried layer leading to a petrosalic classification yet the salt stage class appears to be based on the top very undeveloped soil layer. It was identified earlier as an outlier. Such outliers can have a significant impact on some analyses. In particular, cluster analyses are very sensitive to outliers (Hair et al., 1995).

Hypothesis testing

It may be of interest to test for differences between populations. For example, do turbels have the same mean depth of permafrost as do orthels. An independent two sample *t* test or an ANOVA (for more than 2 groups) can be used if the variable of interest is normally distributed. Ideally the variance of the two samples should be the same, although a separate variance *t* test can be applied at the cost of reduced statistical power if the variance differs. If a number of tests are to be run then the Bonferroni or Dunn-Sidak adjustment must be applied. This also effectively reduces the statistical power, so it is best to identify specific hypotheses of pedological interest rather than applying these techniques indiscriminately. If the data is not normally distributed, either a transformation should be applied or non-parametric tests can be used (Systat, 1997).

Geostatistics

While Antarctic soil literature does associate soil types with timing of glacial activity (which implies some spatial dependence in soil properties), researchers have also found considerable variability in soils over very short distances. For example soils in hollows and depressions had higher concentrations of salts, and a 10-m transect across a "polygon" showed continual variation in soil depth, texture, soil colour, salt and consistence (Campbell and Claridge, 1987). Localised differences in soil moisture will also cause differences in soils (Campbell et al., 1998). The semi-variograms of some of the key soil properties mostly showed no spatial dependence. The only property that did show some spatial pattern was that of chloride, presumably this relates to distance from the coast as discussed earlier.

Larger dataset

Some care is needed in analysing the dataset of the larger McMurdo Dry Valley region as processes in one valley can differ from those in another valley. Salts vary according to origin: in one valley there may be a trapped body of seawater, in another katabatic winds may be stronger depositing more plateau snow (and salts). Glaciers in the McMurdo region have been found to be retreating when others are stationary or advancing (Chinn in Campbell and Claridge, 1987). This lack of synchronicity complicates extrapolation between valleys. Similarity between valleys could be tested by applying the discriminant analysis results to the wider dataset to see if predictive accuracy is maintained.

There is another soils dataset of pit observations collected by Campbell and Claridge. This is maintained by Robert Gibb of Landcare Research. A brief inspection of this shows that this dataset and the Bockheim dataset are in different projections – one would need to be converted. Data observations include the major anions and cations, pH and electrical conductivity (all of which are also in the Bockheim dataset). There is also detailed particle size and mineralogy data, both of which are weak or non-existent in the Bockheim dataset. It is not known if Campbell and Claridge collected any morphological and site information comparable with that in the Bockheim dataset. Due to the mismatch in field observations, statistical analyses on a combined dataset could only be applied to the chemical observations and silt/clay content.

Fit with S-map

S-map is a new spatial soil database for New Zealand that is currently being prototyped. S-map is based on the New Zealand Classification (NZSC) system, which differs from the US Taxonomic system. The NZSC does not currently cater for Gelisols, although it could be relatively easily extended to do so. The key properties in the Dry Valley soils relate to salts and depth of permafrost. Consequently, they wouldn't fit easily into the S-map design. However, just as a solution must be found for Organic soils which have a unique classification structure, no doubt Gelisols could also be incorporated.

Conclusion

Two Bockheim datasets were investigated: one was of the Wright Valley only, the other of the McMurdo Dry Valley region. Both datasets required some manipulation to combine shapefile attributes, extract data from the pit description file, and to tidy non-numeric or missing data. Once this was done both sets are suitable for statistical analyses. The Wright Valley dataset showed that there was considerable variability in key soil properties, which in many cases was not well explained by topographic zone (ecoclim), soil age, taxonomic classification, or spatial dependence.

Preliminary multivariate analyses did appear to identify patterns in the grouping of soils. Further experimentation with alternative linkages and distance metrics, and a more careful selection of independent variates in the cluster analysis might be useful. Similarly, more careful data preparation where taxonomic classes with just a few pits are removed might yield more interesting results by the machine learning algorithms, especially the Ridor and J48 algorithms. If the discriminant analysis is to be taken further then the assumptions should be checked. The next stage in this work is to obtain some pedological input and refine the analyses accordingly.

Few of these statistical analyses can minimise the effect of missing data – most just drop cases with even just one missing item. Consequently, it is preferable that future pedological field trips collect as many of Bockheim's data items as is practical.

References

Bockheim, J. G. (1979) Relative age and origin of soils in the eastern Wright Valley, Antarctica. *Soil Science,* **128** (3) 142-152.

Bockheim, J. G. (2002) Landform and soil development in the McMurdo Dry Valleys, Antarctica: a regional synthesis. *Arctic, antarctic and alpine research,* **34** (3) 308-317.

Campbell, I. B. and Claridge, G. G. C. (1987) *Antarctica: soils, weathering processes and environment,* Elsevier, Amsterdam.

Campbell, I. B., Claridge, G. G. C., Campbell, D. I. and Balks, M. R. (1998) The soil environment of the McMurdo Dry Valleys, Antarctica. In *Ecosystem dynamics in a polar desert. The McMurdo Dry Valleys, Antarctica* (Ed, Priscu, J.) American Geophysical Union, Washington, Antarctic Research Series Vol 72, pp. 297-322.

Hair, J. F., Anderson, R. E., Tatham, R. L. and Black, W. C. (1995) *Multivariate data analysis: with readings,* Prentice-Hall, New Jersey.

Hall, B. and Denton, G. (2005) Surficial geology and geomorphology of eastern and central Wright Valley, Antarctica. *Geomorphology,* **64** 25-65.

McLeod, M. (2004) Soil Scientist, Landcare Research. pp. Email.

Minitab (2003) Minitab 14. StatGuide. Minitab Inc.

New Zealand Antarctic Institute (2001) Ross Sea Region: a state of the environment report for the Ross Sea Region of Antarctica. New Zealand Antarctic Institute, Christchurch.

Soil Survey Staff (1999) *Soil Taxonomy A basic system of soil classification for making and interpreting soil surveys. Natural Resources Conservation Service Agriculture Handbook No. 436,* USDA, Washington.

Systat (1997) Systat 7.0 Statistics. SPSS Inc, Chicago, IL.

Tedrow, J. C. F. and Ugolini, F. C. (1966) Antarctic soils. In *Antarctic soils and soil forming processes* (Ed, Tedrow, J. F. C.) American Geophysical Union, Antarctic Research Series Vol 8, pp. 161-177.

Witten, I. H. and Frank, E. (2000) *Data mining: practical machine learning tools with Java implementations,* Morgan Kaufman, San Franscisco.

Appendix A: ML results run on Weka 15/2/2005. By Brent Martin.

Several ML algorithms were run over the data to try to predict the "subgroup" attribute. Algorithms were chosen for excellent predictive ability, with those producing human-readable rules preferred. Accuracy was measured by stratified (keep class proportions constant) 10-fold cross-validation.

Note: the data set used contained all attributes: removing the saltpan attributes was significantly detrimental to all classifiers (NNGE was able to perform reasonably well in this case if set to generalise aggressively). Replacing the missing saltpan depth with a large number (500) reduced accuracy, so this attribute was used with missing values and the switch was retained.

The resulting accuracies were: (in decreasing order)

- RIDOR: 65.54% (folds=5,use most popular class – ACCH – as default)
- NNGE: 63.27 (gen=2, folds=20)
- J48: 63.27 (minObjs=3, subtree raising=F)
- Naiive Bayes = 56.5
- KNN=54.8 (K=1, best result)
- Base accuracy: 48.6% (assume class=ACCH)

Note: RIDOR is biased towards problems where one class is much more prevalent than the others, as is the case here.

These results are not overly good, considering a guess (class=ACCH) yields nearly 50%.

The following sections give details of the results for each system.

Ridor (65.54%)

Builds an "exception tree" – assumes the most popular class is the correct one, then finds exceptions to this rule.

```
Rules:
subGroup = ACCH   (177.0/91.0)
           Except (SaltpanDepthCM > 13.5) and (SaltpanDepthCM <=
34.5) => subGroup = ACCI  (32.0/0.0) [8.0/0.0]
                    Except (SaltStage <= 5.5) and (SaltpanDepthCM <=
24) => subGroup = ACCF  (13.0/0.0) [3.0/0.0]
           Except (IceCementDepth <= 54) => subGroup = ABCH
(14.0/0.0) [3.0/1.0]
                    Except (IceCementDepth <= 22.5) => subGroup =
ABCB  (6.0/0.0) [1.0/0.0]

=== Summary ===

Correctly Classified Instances         116               65.5367 %
Incorrectly Classified Instances        61               34.4633 %
Kappa statistic                          0.4688
Mean absolute error                      0.0627
Root mean squared error                  0.2503
Relative absolute error                 47.7475 %
Root relative squared error             98.472  %
Total Number of Instances              177
```

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
0	0	0	0	0	ABCA
0.429	0.012	0.6	0.429	0.5	ABCB
0	0	0	0	0	ABCC
0.125	0.012	0.333	0.125	0.182	ABCF
0.115	0.046	0.3	0.115	0.167	ABCH
0	0	0	0	0	ACCA
0	0	0	0	0	ACCD
0	0	0	0	0	ACCE
0.476	0.051	0.556	0.476	0.513	ACCF
0.942	0.352	0.717	0.942	0.814	ACCH
0.857	0.064	0.643	0.857	0.735	ACCI

=== Confusion Matrix ===

a	b	c	d	e	f	g	h	i	j	k	<-- classified as
0	0	0	0	1	0	0	0	0	0	0	a = ABCA
0	3	0	0	3	0	0	0	0	1	0	b = ABCB
0	0	0	0	0	0	0	0	0	0	1	c = ABCC
0	0	0	1	0	0	0	0	3	3	1	d = ABCF
0	1	0	0	3	0	0	0	1	21	0	e = ABCH
0	0	0	0	2	0	0	0	0	0	0	f = ACCA
0	1	0	0	0	0	0	0	1	0	0	g = ACCD
0	0	0	0	0	0	0	0	1	0	1	h = ACCE
0	0	0	0	0	0	0	0	10	5	6	i = ACCF
0	0	0	2	1	0	0	0	1	81	1	j = ACCH
0	0	0	0	0	0	0	0	1	2	18	k = ACCI

J48 Decision Tree builder (63.27%)

Builds a decision tree by finding the most important attribute and splitting the data according to its value, then recursively building a subtree for each value. Prunes the tree to account for noise.

```
J48 pruned tree
------------------

SaltPanSwitch = 0
|   Coherence <= 10
|   |   WeathStage <= 1: ABCB (7.64/1.64)
|   |   WeathStage > 1: ABCH (4.36/2.0)
|   Coherence > 10
|   |   IceCementDepth <= 53
|   |   |   Ghosts <= 5: ABCH (4.0/1.0)
|   |   |   Ghosts > 5: ACCA (3.0/2.0)
|   |   IceCementDepth > 53: ACCH (78.0/16.0)
SaltPanSwitch = 1
|   SaltpanDepthCM <= 13
|   |   MaxCDE <= 18
|   |   |   MaxCDE <= 8
|   |   |   |   WeathStage <= 5: ABCF (4.0)
|   |   |   |   WeathStage > 5: ACCH (4.0/1.0)
|   |   |   MaxCDE > 8: ACCH (24.0/6.0)
```

```
|    |   MaxCDE > 18: ABCH (4.0)
|   SaltpanDepthCM > 13
|    |   SaltStage <= 5
|    |    |   Ghosts <= 8: ACCI (5.0/2.0)
|    |    |   Ghosts > 8
|    |    |    |   Coherence <= 67: ACCE (4.0/2.0)
|    |    |    |   Coherence > 67
|    |    |    |    |   Staining <= 33: ABCF (4.0/2.0)
|    |    |    |    |   Staining > 33: ACCF (13.0/1.0)
|    |   SaltStage > 5: ACCI (18.0/2.0)

Correctly Classified Instances         112               63.2768 %
Incorrectly Classified Instances        65               36.7232 %
Kappa statistic                          0.4601
Mean absolute error                      0.0804
Root mean squared error                  0.2295
Relative absolute error                 61.3024 %
Root relative squared error             90.2769 %
Total Number of Instances              177

=== Detailed Accuracy By Class ===

TP Rate   FP Rate   Precision   Recall  F-Measure   Class
  0         0.006     0          0         0        ABCA
  0.571     0.024     0.5        0.571     0.533    ABCB
  0         0         0          0         0        ABCC
  0.25      0.024     0.333      0.25      0.286    ABCF
  0.154     0.093     0.222      0.154     0.182    ABCH
  0         0.011     0          0         0        ACCA
  0         0         0          0         0        ACCD
  0         0.011     0          0         0        ACCE
  0.476     0.064     0.5        0.476     0.488    ACCF
  0.872     0.275     0.75       0.872     0.806    ACCH
  0.81      0.019     0.85       0.81      0.829    ACCI

=== Confusion Matrix ===

  a  b  c  d  e  f  g  h  i  j  k   <-- classified as
  0  0  0  0  0  0  0  0  0  1  0 |  a = ABCA
  0  4  0  0  3  0  0  0  0  0  0 |  b = ABCB
  0  0  0  1  0  0  0  0  0  0  0 |  c = ABCC
  0  0  0  2  1  0  0  1  2  2  0 |  d = ABCF
  1  3  0  0  4  0  0  0  1 16  1 |  e = ABCH
  0  0  0  0  1  0  0  0  0  1  0 |  f = ACCA
  0  1  0  0  0  0  0  0  1  0  0 |  g = ACCD
  0  0  0  1  0  0  0  0  1  0  0 |  h = ACCE
  0  0  0  2  3  0  0  0 10  4  2 |  i = ACCF
  0  0  0  0  6  2  0  1  2 75  0 |  j = ACCH
  0  0  0  0  0  0  0  0  3  1 17 |  k = ACCI
```

NNGE classifier (nearest neighbour with generalised exemplars – 63.27%)

This classifier forms rules over arbitrary subsets of the data space. Future data points falling within these rules are classified according to the rule. Those falling outside are classified using nearest neighbour. The "rules" indicate the range of values each attribute must fall in. The number after each rule indicates how many data points are inside the rule.

Rules generated

class ABCB IF

- Staining=0.0 ^ 0.0<=Coherence<=10.0 ^ 0.0<=Vissalt<=60.0 ^ Ghosts=0.0 ^ 8.0<=IceCementDepth<=63.0 ^ 0.0<=MaxCDE<=9.0 ^ SaltStage=1.0 ^ SaltpanDepthCM=0.0 ^ SaltPanSwitch in {0} ^ WeathStage=1.0 ^ 0.0<=SBF314<=248.0 ^ 0.0<=MaxECdSm<=1.6 ^ 0.0<=SaltsTo70cm<=456.0 ^ EcoClimate in {InlandValleyFloor} (3)

- Staining=0.0 ^ 8.0<=Coherence<=9.0 ^ Vissalt=0.0 ^ Ghosts=0.0 ^ 8.0<=IceCementDepth<=9.0 ^ 0.0<=MaxCDE<=9.0 ^ SaltStage=0.0 ^ SaltpanDepthCM=0.0 ^ SaltPanSwitch in {0} ^ WeathStage=1.0 ^ 0.0<=SBF314<=540.0 ^ 0.37<=MaxECdSm<=1.1 ^ 65.0<=SaltsTo70cm<=354.0 ^ EcoClimate in {InlandValleySide} (3)

class ABCF IF

- 17.0<=Staining<=30.0 ^ 83.0<=Coherence<=100.0 ^ 40.0<=Vissalt<=48.0 ^ 10.0<=Ghosts<=13.0 ^ 83.0<=IceCementDepth<=120.0 ^ 12.0<=MaxCDE<=18.0 ^ SaltStage=5.0 ^ 17.0<=SaltpanDepthCM<=27.0 ^ SaltPanSwitch in {1} ^ 5.0<=WeathStage<=6.0 ^ SBF314=0.0 ^ 0.0<=MaxECdSm<=48.0 ^ 0.0<=SaltsTo70cm<=9185.0 ^ EcoClimate in {InlandValleySide} (2)

- 23.0<=Staining<=36.0 ^ 32.0<=Coherence<=100.0 ^ 11.0<=Vissalt<=44.0 ^ 5.0<=Ghosts<=25.0 ^ 80.0<=IceCementDepth<=100.0 ^ 4.0<=MaxCDE<=9.0 ^ 4.0<=SaltStage<=5.0 ^ 5.0<=SaltpanDepthCM<=14.0 ^ SaltPanSwitch in {1} ^ 4.0<=WeathStage<=5.0 ^ 0.0<=SBF314<=109.0 ^ 50.0<=MaxECdSm<=99.0 ^ 3770.0<=SaltsTo70cm<=9327.0 ^ EcoClimate in {InlandValleyFloor} (5)

class ABCH IF

- 0.0<=Staining<=6.0 ^ 0.0<=Coherence<=49.0 ^ 0.0<=Vissalt<=80.0 ^ 0.0<=Ghosts<=8.0 ^ 25.0<=IceCementDepth<=110.0 ^ 9.0<=MaxCDE<=12.0 ^ 0.0<=SaltStage<=4.0 ^ SaltpanDepthCM=0.0 ^ SaltPanSwitch in {0} ^ 2.0<=WeathStage<=3.0 ^ 0.0<=SBF314<=580.0 ^ 0.0<=MaxECdSm<=3.6 ^ 0.0<=SaltsTo70cm<=806.0 ^ EcoClimate in {InlandValleyFloor,InlandValleySide} (5)

- 12.0<=Staining<=90.0 ^ 52.0<=Coherence<=105.0 ^ 12.0<=Vissalt<=20.0 ^ 0.0<=Ghosts<=17.0 ^ 52.0<=IceCementDepth<=105.0 ^ 24.0<=MaxCDE<=40.0 ^ 4.0<=SaltStage<=5.0 ^ 0.0<=SaltpanDepthCM<=14.0 ^ SaltPanSwitch in {0,1} ^ 3.0<=WeathStage<=6.0 ^ 0.0<=SBF314<=860.0 ^ 0.0<=MaxECdSm<=7.5 ^ 0.0<=SaltsTo70cm<=1648.0 ^ EcoClimate in {InlandValleyFloor,UplandValley} (5)

- 13.0<=Staining<=31.0 ^ 53.0<=Coherence<=77.0 ^ 0.0<=Vissalt<=25.0 ^ 0.0<=Ghosts<=9.0 ^ 53.0<=IceCementDepth<=77.0 ^ 12.0<=MaxCDE<=18.0 ^ SaltStage=1.0 ^ SaltpanDepthCM=0.0 ^ SaltPanSwitch in {0} ^ WeathStage=4.0 ^ 0.0<=SBF314<=29.0 ^ MaxECdSm=0.0 ^ SaltsTo70cm=0.0 ^ EcoClimate in {UplandValley} (3)

- 32.0<=Staining<=65.0 ^ 55.0<=Coherence<=110.0 ^ 0.0<=Vissalt<=35.0 ^ 10.0<=Ghosts<=21.0 ^ 60.0<=IceCementDepth<=122.0 ^ 18.0<=MaxCDE<=24.0 ^ 1.0<=SaltStage<=2.0 ^ SaltpanDepthCM=0.0 ^ SaltPanSwitch in {0} ^ WeathStage=3.0 ^ 0.0<=SBF314<=125.0 ^ MaxECdSm=0.0 ^ SaltsTo70cm=0.0 ^ EcoClimate in {InlandValleyFloor} (3)

- 33.0<=Staining<=65.0 ^ 8.0<=Coherence<=33.0 ^ Vissalt=60.0 ^ 11.0<=Ghosts<=13.0 ^ 65.0<=IceCementDepth<=103.0 ^ 16.0<=MaxCDE<=32.0 ^ SaltStage=4.0 ^ 5.0<=SaltpanDepthCM<=6.0 ^ SaltPanSwitch in {1} ^ WeathStage=5.0 ^ 34.0<=SBF314<=122.0 ^ 13.0<=MaxECdSm<=15.5 ^ 1027.0<=SaltsTo70cm<=1547.0 ^ EcoClimate in {UplandValley} (2)

- Staining=0.0 ^ 6.0<=Coherence<=65.0 ^ Vissalt=0.0 ^ Ghosts=0.0 ^ 52.0<=IceCementDepth<=65.0 ^ MaxCDE=12.0 ^ SaltStage=0.0 ^ SaltpanDepthCM=0.0 ^ SaltPanSwitch in {0} ^ WeathStage=0.0 ^ SBF314=0.0 ^ 0.0<=MaxECdSm<=4.6 ^ 0.0<=SaltsTo70cm<=417.0 ^ EcoClimate in {InlandValleyFloor} (2)

class ACCA IF

- 0.0<=Staining<=16.0 ^ 12.0<=Coherence<=26.0 ^ Vissalt=0.0 ^ 0.0<=Ghosts<=26.0 ^ 7.0<=IceCementDepth<=27.0 ^ 12.0<=MaxCDE<=18.0 ^ 0.0<=SaltStage<=1.0 ^ SaltpanDepthCM=0.0 ^ SaltPanSwitch in {0,1} ^ WeathStage=1.0 ^ SBF314=0.0 ^ MaxECdSm=0.0 ^ SaltsTo70cm=0.0 ^ EcoClimate in {UplandValley} (2)

class ACCF IF

- 15.0<=Staining<=80.0 ^ 28.0<=Coherence<=108.0 ^ 28.0<=Vissalt<=93.0 ^ 14.0<=Ghosts<=26.0 ^ 80.0<=IceCementDepth<=120.0 ^ 12.0<=MaxCDE<=24.0 ^ 4.0<=SaltStage<=5.0 ^ 10.0<=SaltpanDepthCM<=17.5 ^ SaltPanSwitch in {1} ^ 4.0<=WeathStage<=5.0 ^ 0.0<=SBF314<=22.0 ^ 0.0<=MaxECdSm<=50.0 ^ 0.0<=SaltsTo70cm<=5466.0 ^ EcoClimate in {InlandValleyFloor,InlandValleySide} (5)

- 35.0<=Staining<=110.0 ^ 70.0<=Coherence<=110.0 ^ 35.0<=Vissalt<=110.0 ^ 11.0<=Ghosts<=26.0 ^ 70.0<=IceCementDepth<=110.0 ^ 0.0<=MaxCDE<=18.0 ^ 4.0<=SaltStage<=5.0 ^ 19.0<=SaltpanDepthCM<=40.0 ^ SaltPanSwitch in {1} ^ 4.0<=WeathStage<=6.0 ^ 0.0<=SBF314<=155.0 ^ 0.0<=MaxECdSm<=30.0 ^ 0.0<=SaltsTo70cm<=4698.0 ^ EcoClimate in {InlandValleyFloor,InlandValleySide} (7)

- 90.0<=Staining<=92.0 ^ 90.0<=Coherence<=92.0 ^ 45.0<=Vissalt<=47.0 ^ 9.0<=Ghosts<=16.0 ^ 90.0<=IceCementDepth<=92.0 ^ 6.0<=MaxCDE<=8.0 ^ 4.0<=SaltStage<=5.0 ^ 10.0<=SaltpanDepthCM<=17.0 ^ SaltPanSwitch in {1} ^ WeathStage=6.0 ^ 2.0<=SBF314<=210.0 ^ 50.0<=MaxECdSm<=55.0 ^ 7198.0<=SaltsTo70cm<=7681.0 ^ EcoClimate in {InlandValleySide} (3)

class ACCH IF

- 0.0<=Staining<=100.0 ^ 18.0<=Coherence<=110.0 ^ 0.0<=Vissalt<=75.0 ^ 0.0<=Ghosts<=13.0 ^ 0.0<=IceCementDepth<=120.0 ^ 8.0<=MaxCDE<=18.0 ^ 4.0<=SaltStage<=6.0 ^ 7.0<=SaltpanDepthCM<=13.0 ^ SaltPanSwitch in {1} ^ 4.0<=WeathStage<=6.0 ^ 0.0<=SBF314<=254.0 ^ 0.0<=MaxECdSm<=2.9 ^ 0.0<=SaltsTo70cm<=426.0 ^ EcoClimate in {InlandValleyFloor,InlandValleySide,UplandValley} (10)

- 10.0<=Staining<=14.0 ^ 104.0<=Coherence<=115.0 ^ 0.0<=Vissalt<=10.0 ^ 0.0<=Ghosts<=14.0 ^ 105.0<=IceCementDepth<=115.0 ^ 9.0<=MaxCDE<=12.0 ^ 1.0<=SaltStage<=2.0 ^ SaltpanDepthCM=0.0 ^ SaltPanSwitch in {0} ^ WeathStage=3.0 ^ 0.0<=SBF314<=255.0 ^ 0.0<=MaxECdSm<=6.9 ^ 0.0<=SaltsTo70cm<=1833.0 ^ EcoClimate in {InlandValleySide} (3)

- 20.0<=Staining<=21.0 ^ 80.0<=Coherence<=100.0 ^ 10.0<=Vissalt<=20.0 ^ 0.0<=Ghosts<=6.0 ^ 80.0<=IceCementDepth<=100.0 ^ MaxCDE=18.0 ^ 5.0<=SaltStage<=6.0 ^ SaltpanDepthCM=4.0 ^ SaltPanSwitch in {1} ^ 5.0<=WeathStage<=6.0 ^ SBF314=0.0 ^ MaxECdSm=0.0 ^ SaltsTo70cm=0.0 ^ EcoClimate in {UplandValley} (2)

- 23.0<=Staining<=40.0 ^ 55.0<=Coherence<=60.0 ^ 3.0<=Vissalt<=22.0 ^ 10.0<=Ghosts<=20.0 ^ 55.0<=IceCementDepth<=60.0 ^ 24.0<=MaxCDE<=30.0 ^ SaltStage=1.0 ^ SaltpanDepthCM=0.0 ^ SaltPanSwitch in {0} ^ 2.0<=WeathStage<=3.0 ^ SBF314=0.0 ^ MaxECdSm=0.0 ^ SaltsTo70cm=0.0 ^ EcoClimate in {InlandValleySide,UplandValley} (2)

- 52.0<=Staining<=77.0 ^ 110.0<=Coherence<=120.0 ^ 39.0<=Vissalt<=100.0 ^ 16.0<=Ghosts<=40.0 ^ 110.0<=IceCementDepth<=120.0 ^ 12.0<=MaxCDE<=18.0 ^ 4.0<=SaltStage<=5.0 ^ 10.0<=SaltpanDepthCM<=11.0 ^ SaltPanSwitch in {1} ^ WeathStage=6.0 ^ 1.0<=SBF314<=50.0 ^ MaxECdSm=0.0 ^ SaltsTo70cm=0.0 ^ EcoClimate in {InlandValleySide} (3)

- 55.0<=Staining<=85.0 ^ 63.0<=Coherence<=100.0 ^ 29.0<=Vissalt<=55.0 ^ 25.0<=Ghosts<=40.0 ^ 63.0<=IceCementDepth<=100.0 ^ 12.0<=MaxCDE<=18.0 ^ 3.0<=SaltStage<=4.0 ^ SaltpanDepthCM=0.0 ^ SaltPanSwitch in {0} ^ 4.0<=WeathStage<=6.0 ^ 0.0<=SBF314<=224.0 ^ MaxECdSm=0.0 ^ SaltsTo70cm=0.0 ^ EcoClimate in {InlandValleySide} (3)

- 63.0<=Staining<=115.0 ^ 63.0<=Coherence<=130.0 ^ 12.0<=Vissalt<=63.0 ^ 4.0<=Ghosts<=47.0 ^ 100.0<=IceCementDepth<=115.0 ^ MaxCDE=9.0 ^ 4.0<=SaltStage<=6.0 ^ SaltpanDepthCM=12.0 ^ SaltPanSwitch in {1} ^ WeathStage=6.0 ^ SBF314=5.0 ^ 55.0<=MaxECdSm<=75.0 ^ 7157.0<=SaltsTo70cm<=9963.0 ^ EcoClimate in {InlandValleySide} (2)

- 8.0<=Staining<=100.0 ^ 15.0<=Coherence<=115.0 ^ 11.0<=Vissalt<=95.0 ^ 0.0<=Ghosts<=45.0 ^ 70.0<=IceCementDepth<=115.0 ^ 8.0<=MaxCDE<=18.0 ^ 1.0<=SaltStage<=2.0 ^ SaltpanDepthCM=0.0 ^ SaltPanSwitch in {0} ^ 2.0<=WeathStage<=6.0 ^ 0.0<=SBF314<=300.0 ^ 0.0<=MaxECdSm<=18.0 ^ 0.0<=SaltsTo70cm<=4353.0 ^ EcoClimate in {InlandValleySide} (22)

- 8.0<=Staining<=15.0 ^ 11.0<=Coherence<=24.0 ^ Vissalt=0.0 ^ Ghosts=0.0 ^ 75.0<=IceCementDepth<=105.0 ^ 8.0<=MaxCDE<=12.0 ^ SaltStage=1.0 ^ SaltpanDepthCM=0.0 ^ SaltPanSwitch in {0} ^

2.0<=WeathStage<=3.0 ^ 0.0<=SBF314<=345.0 ^ 0.0<=MaxECdSm<=11.0 ^ 0.0<=SaltsTo70cm<=1603.0 ^ EcoClimate in {InlandValleySide} (3)

- 9.0<=Staining<=43.0 ^ 9.0<=Coherence<=105.0 ^ 0.0<=Vissalt<=105.0 ^ 0.0<=Ghosts<=40.0 ^ 65.0<=IceCementDepth<=115.0 ^ 8.0<=MaxCDE<=12.0 ^ 1.0<=SaltStage<=4.0 ^ 0.0<=SaltpanDepthCM<=6.0 ^ SaltPanSwitch in {0,1} ^ 1.0<=WeathStage<=4.0 ^ 0.0<=SBF314<=163.0 ^ 0.0<=MaxECdSm<=19.0 ^ 0.0<=SaltsTo70cm<=3377.0 ^ EcoClimate in {InlandValleyFloor} (20)

class ACCI IF

- 0.0<=Staining<=100.0 ^ 100.0<=Coherence<=120.0 ^ 28.0<=Vissalt<=65.0 ^ 14.0<=Ghosts<=38.0 ^ 100.0<=IceCementDepth<=120.0 ^ 9.0<=MaxCDE<=10.5 ^ SaltStage=6.0 ^ 14.0<=SaltpanDepthCM<=36.0 ^ SaltPanSwitch in {1} ^ WeathStage=6.0 ^ 0.0<=SBF314<=2.0 ^ 65.0<=MaxECdSm<=80.0 ^ 7607.0<=SaltsTo70cm<=13116.0 ^ EcoClimate in {InlandValleySide} (2)

- 0.0<=Staining<=100.0 ^ 15.0<=Coherence<=100.0 ^ 0.0<=Vissalt<=58.0 ^ 0.0<=Ghosts<=6.0 ^ 0.0<=IceCementDepth<=70.0 ^ 9.0<=MaxCDE<=18.0 ^ 5.0<=SaltStage<=6.0 ^ 15.0<=SaltpanDepthCM<=28.0 ^ SaltPanSwitch in {1} ^ 4.0<=WeathStage<=6.0 ^ 0.0<=SBF314<=429.0 ^ 0.0<=MaxECdSm<=55.0 ^ 0.0<=SaltsTo70cm<=4182.0 ^ EcoClimate in {InlandValleyFloor,InlandValleySide} (4)

- 33.0<=Staining<=105.0 ^ 58.0<=Coherence<=105.0 ^ 28.0<=Vissalt<=103.0 ^ 0.0<=Ghosts<=50.0 ^ 80.0<=IceCementDepth<=105.0 ^ MaxCDE=12.0 ^ SaltStage=6.0 ^ 14.0<=SaltpanDepthCM<=34.0 ^ SaltPanSwitch in {1} ^ WeathStage=6.0 ^ 0.0<=SBF314<=301.0 ^ 0.0<=MaxECdSm<=85.0 ^ 0.0<=SaltsTo70cm<=10101.0 ^ EcoClimate in {InlandValleyFloor,InlandValleySide} (11)

=== Summary ===

Correctly Classified Instances	112	63.2768 %
Incorrectly Classified Instances	65	36.7232 %
Kappa statistic	0.4462	
Mean absolute error	0.0668	
Root mean squared error	0.2584	
Relative absolute error	50.8785 %	
Root relative squared error	101.6494 %	
Total Number of Instances	177	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
0	0	0	0	0	ABCA
0.286	0.018	0.4	0.286	0.333	ABCB
0	0	0	0	0	ABCC
0.125	0	1	0.125	0.222	ABCF
0.346	0.093	0.391	0.346	0.367	ABCH
0	0	0	0	0	ACCA
0	0.006	0	0	0	ACCD
0	0	0	0	0	ACCE
0.619	0.064	0.565	0.619	0.591	ACCF

```
  0.849      0.352       0.695      0.849      0.764    ACCH
  0.667      0.032       0.737      0.667      0.7      ACCI

=== Confusion Matrix ===

  a  b  c  d  e  f  g  h  i  j  k   <-- classified as
  0  0  0  0  1  0  0  0  0  0  0 |  a = ABCA
  0  2  0  0  3  0  1  0  0  1  0 |  b = ABCB
  0  0  0  0  1  0  0  0  0  0  0 |  c = ABCC
  0  0  0  1  0  0  0  0  2  4  1 |  d = ABCF
  0  1  0  0  9  0  0  0  0 16  0 |  e = ABCH
  0  0  0  0  0  0  0  0  0  2  0 |  f = ACCA
  0  1  0  0  0  0  0  0  0  1  0 |  g = ACCD
  0  0  0  0  0  0  0  0  2  0  0 |  h = ACCE
  0  0  0  0  0  0  0  0 13  5  3 |  i = ACCF
  0  1  0  0  7  0  0  0  4 73  1 |  j = ACCH
  0  0  0  0  2  0  0  0  2  3 14 |  k = ACCI
```